

AD-A098 625

ROWLAND AND CO. HADDONFIELD N J

F/G 5/9

TRAINING IMPLICATIONS OF AIRBORNE APPLICATIONS OF AUTOMATED SPE--ETC(U)

OCT 80 P E VAN HEMEL, S B VAN HEMEL, W J KING N61339-80-D-0009

UNCLASSIFIED

NAVTRAEQUIPC-80-D-0009-015 NL

101
20-40-10

101
20-40-10

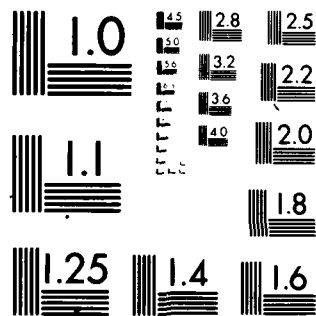
END

DATE

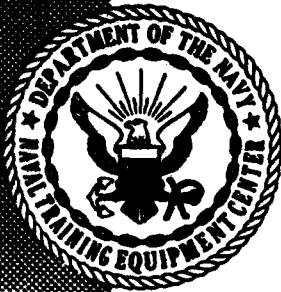
FILED

8-8-81

DTIC



MICROCOPY RESOLUTION TEST CHART
 NATIONAL BUREAU OF STANDARDS-1963-A



AD A098625

FILE COPY

LEVEL II (2) 1 fw
Technical Report NAVTRAEQUIPCEN 80-D-0009-0155-1

TRAINING IMPLICATIONS OF AIRBORNE APPLICATIONS
OF AUTOMATED SPEECH RECOGNITION TECHNOLOGY

Paul E. Van Hemel, Susan B. Van Hemel,
W. Judson King
Ergonomics Associates, Inc.
999 Woodcock Road, P. O. Box 20987
Orlando, Florida 32814

and

Robert Breaux
Human Factors Laboratory
Naval Training Equipment Center
Orlando, Florida 32813

FINAL TECHNICAL REPORT--July 1980-October 1980

DTIC
ELECTE
S MAY 8 1981 D

A

DoD DISTRIBUTION STATEMENT

Approved for public release;
distribution unlimited.

NAVAL TRAINING EQUIPMENT CENTER
ORLANDO, FLORIDA 32813

81 5 07 017

NAVTRACQUIPCEN 88-D-0009-0155-1

GOVERNMENT RIGHTS IN DATA STATEMENT

Reproduction of this publication in whole or in part is permitted for any purpose of the United States Government.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAVTRAEQUIPCEN 88-D-0009-0155-1	2. GOVT ACCESSION NO. AD-A098625	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) TRAINING IMPLICATIONS OF AIRBORNE APPLICATIONS OF AUTOMATED SPEECH RECOGNITION TECHNOLOGY.		5. TYPE OF REPORT & PERIOD COVERED Final Technical Report July 1980 - October 1980
7. AUTHOR(s) Paul E./Van Hemel, Susan B./Van Hemel, W. Judson/King, Robert/Breaux		8. CONTRACT OR GRANT NUMBER(s) N61339-80-D-0009-0155
9. PERFORMING ORGANIZATION NAME AND ADDRESS Ergonomics Associates, Inc. ¹ 999 Woodcock Road, P.O. Box 20987 Orlando, Florida 32814		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS 0155-1P1
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Training Equipment Center, Code N-82 Orlando, Florida 32813		12. REPORT DATE
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		13. NUMBER OF PAGES 59
		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION DOWNGRADING SCHEDULE
16. DISTRIBUTION STATEMENT (of this Report) Approved for public release; distribution unlimited.		
17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)		
18. SUPPLEMENTARY NOTES ¹ Performed as Task 1P1 under subcontract to: Rowland and Company P.O. Box 61 Haddonfield, New Jersey 08033		
19. KEY WORDS (Continue on reverse side if necessary and identify by block number) VTAG Instructional Systems Development Applied Voice Technology Media Selection Automated Speech Recognition Automated Training Human-Machine Communication Human Factors (A-R)		
20. ABSTRACT (Continue on reverse side if necessary and identify by block number) Developments in automated voice recognition and synthesis have made feasible the implementation of automated speech recognition technology in airborne systems. Research systems using voice technology at the Naval Training Equipment Center and at the Naval Air Development Center were analyzed to determine the human factors of using automated speech recognition for communication with machines. The human factors identified present some unique training implications. The following specific recommendations were made with respect to Instructional Systems Development and particularly to the development of training media.		

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

Instructional systems using ~~Automated Speech Recognition~~ ^{g g} (ASR) should:

- (1) be developed by instructional designers who have had hands-on experience with ASR technology;
- (2) provide ASR speech behavior models, especially correct ones, for trainees to emulate, with the models chosen to illustrate specific factors in achieving successful recognition;
- (3) provide convenient and effective means by which trainees can evaluate their own speech behavior; *7, d*
- (4) provide convenient voice recognition test and voice reference pattern update capabilities under trainee control.

Recommendations are also provided for

Training for the use of airborne ASR systems, should:

1. be based on front-end analyses which are performed by professionals who thoroughly understand the human factors of human-ASR interaction;
2. be based on front-end analyses which explicitly consider the integration of ASR into airborne task performance;
3. prepare users of speaker-dependent ASR systems to register voice reference patterns in effective physical and psychological context, which may or may not be done in the training setting;
4. encourage trainees to experiment with ASR use and to develop a personal style of information exchange which optimizes their task performance;
5. include specific instruction and practice in dealing with recognition failures.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

This analysis effort represents a first step in introducing the technology of Automated Speech Recognition to training system analysts and designers. It is intended to provide a brief background of ASR and a discussion of the training implications that can be expected from the interactions between human speakers and ASR systems.

The Ergonomics/NAVTRAEQUIPCEN study team is grateful to the command and staff of the Naval Air Development Center (NAVAIRDEVCEN), Warminster, Pennsylvania. LT Steven D. Harris and Dr. Norman W. Warner were especially helpful in arranging for hands-on experience with the Voice Recognition and Synthesis (VRAS) system and in providing background on voice technology research at the NAVAIRDEVCEN.

Rhine.

R. BIRD
Analysis Manager

A

TABLE OF CONTENTS

<u>Section</u>	<u>Page</u>
I INTRODUCTION	7
Airborne Automated Speech Recognition Technology	7
Training for Users of Airborne ASR Systems	8
Study Origins, Objectives, and Approach	10
II NAVAIRDEVCCEN AND NAVTRAEQUIPCEN ASR RESEARCH	13
NAVAIRDEVCCEN ASR Research	13
The VRAS System	14
Outlook	15
NAVTRAEQUIPCEN ASR Research	16
Air Controller Training	16
GCA-CTS as an ASR System	17
Outlook	18
III HUMAN FACTORS CONSIDERATIONS IN ASR TRAINING	19
Human-ASR Interactions: General Issues	21
User Resistance to Change	21
User Expectations for Artificially Intelligent Systems	23
Feedback/Verification of Speech Input	24
Human-Machine Competition	26
Non-Conventional User Strategies/Techniques Required	27
Human-ASR Interactions: Speech Control Factors	28
Constraints on User Speech Patterns	28
Voice Reference Pattern Formation	32
Speech Recognition Performance	36
IV GUIDELINES FOR THE USE OF ASR IN TRAINING: OVERCOMING HUMAN FACTORS PROBLEMS	40
Introduction	40
Speech Discipline	41
Principles and Strategies	45

TABLE OF CONTENTS (continued)

<u>Section</u>	<u>Page</u>
V	
IMPACT OF AUTOMATED SPEECH RECOGNITION TECHNOLOGY ON MEDIA SELECTION AND OTHER ISD PROCEDURES	48
Instructional Systems Development for	
Operational ASR	48
Hands-on Media Selection	48
Academic Media Selection	50
Other ISD Procedures	50
ASR as a Training Medium	51
VI	
CONCLUSIONS AND RECOMMENDATIONS	52
General Remarks	52
ASR in Training Systems	52
1. Hands-on ASR	52
2. Speech Behavior Models	53
3. Speech Evaluation and Feedback	53
4. Recognition Test and Voice Reference Pattern Update	53
Training for Operational Use of ASR	53
1. Human Factors Training Analysis	53
2. ASR Integrated in Task Performance	54
3. Personal Style in ASR Use	54
4. Voice Reference Pattern Formation Context	54
5. Recognition Failure Experience	55
A Final Comment	55
REFERENCES	56

LIST OF ILLUSTRATIONS

<u>Figure</u>		<u>Page</u>
1	Schematic Representation of ASR Process . .	9
2	Study Approach	11
3	Human Factors Restrictions on the Automated Speech Recognition Channel . . .	20

SECTION I

INTRODUCTION

AIRBORNE AUTOMATED SPEECH RECOGNITION TECHNOLOGY

Automated Speech Technology is a technology which has been promoted as beneficial if applied to a wide variety of Navy operational and training programs.¹ Caution demands that research must pave the way for airborne applications of Automated Speech Recognition (ASR) technology.² Research programs in military and commercial laboratories have already brought the technology to a level of utility and reliability which is sufficient for some airborne applications, and within a very few years ASR systems will be available which can reliably handle a wide variety of airborne tasks requiring man-machine interactions. The question arises then of the training implications of the use of ASR in airborne systems.

Some of the airborne tasks which ASR systems can perform, or assist in performing, include monitoring system status, activating switches, and adjusting controls. Others include various data handling and transfer tasks, such as presentation of data to the operator upon request, entry of data by the operator, and processing operator requests for various calculations or decision-aiding functions.³

The first application of ASR in Navy aircraft cockpits can be expected to occur within the next two to four years. For example, the Navy is currently exploring the possibility of adding limited ASR capability to the A-7E during the Performance

¹Feuge, R. L. & Geer, C. W. Integrated applications of automated speech technology final report, ONR-CR213-158-1AF. Arlington, VA: Office of Naval Research, 1978.

²Lea, W. A. Critical issues in airborne applications of speech recognition. Los Angeles, CA: Speech Communications Research Laboratory, 1980.

³Curran, M. Voice integrated systems. In R. Breaux, M. Curran, & E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

Enhancement program for that aircraft in FY-81. The system would be for fuel monitoring and fuel consumption calculations, and would utilize speech recognition technology comparable to that now commercially available.

Other aircraft may soon have ASR capability. The Navy is currently assessing crew station workloads for potential ASR application in fighter-attack, patrol, and advanced early warning aircraft. Some probable initial functions for ASR include radio frequency switching and various data entry tasks.

In all of these airborne applications, automated speech recognition would be employed as a channel for man-machine communication which can be used when other channels (e.g., manual, visual) are occupied. The advantages of being able to interact with aircraft systems using the voice mode when hands and eyes are busy are obvious. The judicious employment of ASR promises to ease critical crew workload problems, and to allow aircrews to perform many mission tasks more quickly and with fewer errors than has been possible using conventional systems.

The concept of man-machine interaction using automated speech recognition is simple: the human operator speaks, and the machine understands. To explain it a bit more functionally: the speech pre-processor operates on the speech signals it receives, decides what has been said, and passes on a computer-language translation to the host computer system. The host computer interprets the message, and performs an appropriate response, which may be to set a switch, report on the status of an aircraft system, or merely to verify that the message was received. Often, ASR systems are combined with voice synthesizers to use computer-produced speech to communicate with the human operator. Figure 1 shows a simplified schematic representation of the relationships between these functions. The entire process is accomplished quickly, and when correct recognition occurs the system appears to perform as would a capable listening human, such as a copilot.

TRAINING FOR USERS OF AIRBORNE ASR SYSTEMS

The Navy must develop procedures to train users of airborne ASR systems, because despite similarities between ASR systems and listening humans, talking to a machine is not the same as talking to a copilot. The apparent similarity between ASR and human speech

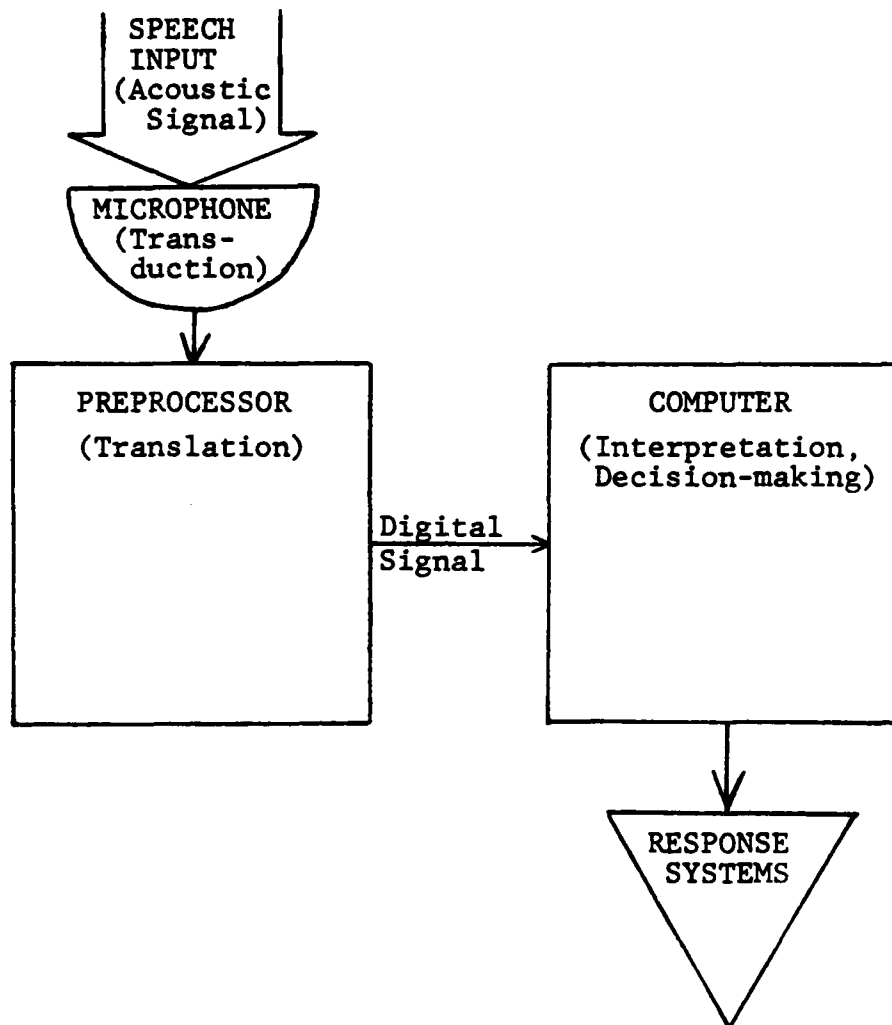


Figure 1. Simplified schematic representation of the Automated Speech Recognition process.

understanding has had significant impact on the training of human operators who use currently available ASR systems. In fact, ASR systems do not perform in exactly the same way as a human listener would. They can understand only limited vocabulary, spoken in certain constrained ways. Nor do they interact with humans quite as other machines do; they allow, and in fact require, a new communications mode.

Thus it will be important to train users of ASR to be cognizant and tolerant of the systems' limitations, as well as to train them to take full advantage of their capabilities. An automated speech understanding system will work best for operators who are aware 1) that it is different from any other machine or human with which they have communicated in the past, and 2) that they will enjoy the full benefits of ASR only if they learn to adapt to its requirements by adjusting their speech patterns. The magnitude of the adjustment required varies with many factors and may not always be extensive.

STUDY ORIGIN, OBJECTIVES, AND APPROACH

The present study is part of a Navy research program to improve training through application of voice technology in self-paced adaptive training systems. The study is to examine the effects of operational applications of voice technology on training system development. The objectives of the present study were to:

- (1) review selected Navy research on airborne and training applications of automated speech recognition technology;
- (2) develop a list of ASR-specific human factors with implications for aircrew training systems;
- (3) determine the implications of ASR-specific human factors for media selection in Instructional Systems Development (ISD).

Figure 2 shows an outline of the approach taken to achieve the objectives of the study. As the figure shows, the ultimate aim of the study was the development of recommendations for ISD procedures to achieve transfer of technology developed in research to application by instructional system designers.

The study began with reviews of research on airborne and training applications of ASR, including

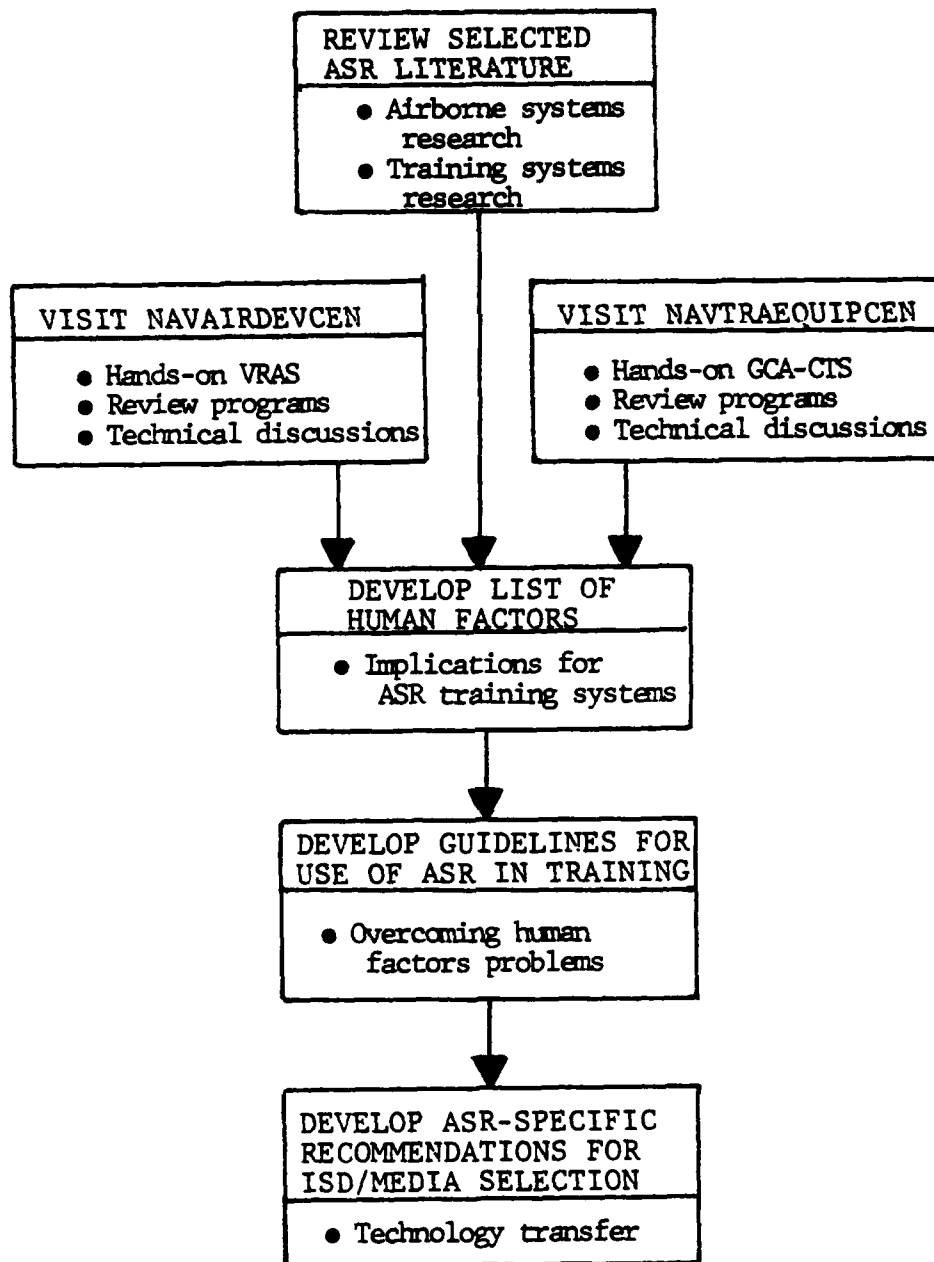


Figure 2. Study approach.

visits to the NAVAIRDEVCEN and the Naval Training Equipment Center (NAVTRAEQUIPCEN) and review of research reports related to their ASR programs. These reviews and hands-on experience with experimental or prototype Navy ASR systems provided the basis for developing a list of human factors with implications for ASR training systems. The list in turn provided a background for an evaluation of the suitability of ISD and other approaches to the development of systems for ASR user training. Recommendations could then be developed for ISD media selection procedures specific to training for systems using ASR technology.

SECTION II

NAVAIRDEVCEN AND NAVTRAEQUIPCEN ASR RESEARCH

The first part of the present study was a review to become familiar with ASR-specific research programs at the NAVAIRDEVCEN and the NAVTRAEQUIPCEN and to obtain hands-on experience with some experimental or prototype ASR systems. The purpose of the review and hands-on experience was to form a basis for development of a list of human factors affecting formation of voice reference patterns, affecting recognition of speech, affecting user acceptance, and therefore affecting criteria used by instructional designers in selecting media in accordance with Section 3.11 of MIL-T-29053A(TD) dated 14 December 1979.

NAVAIRDEVCEN ASR RESEARCH

The research currently in progress at the NAVAIRDEVCEN has two major goals. One is to pursue the development of speech understanding systems and syntactical processors in order to have the best possible systems available for operational use. The other is to study crew tasks and workload to determine the benefits and risks of the potential applications of airborne ASR systems. Together, these two lines of research should be able to:

- 1) identify the specific crewstation applications for which ASR is best suited, and
- 2) develop hardware and software which can effectively aid aircrews in the performance of their tasks.

Lane and Harris concisely explain the philosophy that has guided the NAVAIRDEVCEN effort to ensure that ASR is applied effectively in airborne platforms, as follows:

If voice systems are to be effective in military crewstations, their design must be tailored to the tasks required of a given operator in a specific platform. A thorough analysis of each operator's tasks in a variety of mission contexts must be performed, and the points at which excess workload is occurring must be identified. These overload conditions must be systematically examined for tasks of the type that can be effectively enhanced by

voice input and output...Techniques have been developed for evaluating crew tasks to determine those which might be augmented by voice and to identify crew actions (control actuation, display configuration, data entry, etc.) which could be accomplished through voice commands. For each potential application, tradeoff matrices are constructed which compare probable increases in system performance to factors of technical feasibility, risk, cost and potential interference with other system tasks. (pp. 4-5)⁴

The VRAS System

One of the products of the ASR development program at the NAVAIRDEVCON is the Voice Recognition and Synthesis (VRAS) system, a syntactical processing program which can be adapted to various computers and operating systems. It accepts the output of an isolated-word voice recognition preprocessor, such as Threshold Technology's Threshold 500, and performs semantic and syntactical processing which allows it to interface with various aircraft systems. Thus, VRAS allows an operator to query the status of various subsystems; it interfaces with the system in question, obtains a reading or status indication, and presents the information to the operator through its speech synthesis or CRT readout. When appropriately interfaced, VRAS can also operate on aircraft systems to change settings, immediately or when a stated condition is fulfilled (e.g. "when target distance is less than 5 miles, report it and change guns to armed")^{5,6,7}

⁴Lane, N. E. and Harris, S. D. Conversations with weapon systems: Crewstation applications of interactive voice technology. In Yearbook on Navy manpower, personnel & training research and development. Washington, D.C.: Office of the Chief of Naval Operations, in press.

⁵Streib, M. I. & Stokes, J. M. Military applications of task oriented grammars, Technical Report 1400.10-B. Willow Grove, PA: Analytics, 1980.

⁶Lane & Harris, op. cit.

⁷Streib, M. I. & Preston, J. F. Voice recognition/synthesis for the Advanced Integrated Display System (AIDS), Technical Report 1343. Willow Grove, PA: Analytics, 1978.

VRAS is currently being implemented on NAVAIRDEVCECEN'S Advanced Integrated Display System (AIDS), a cockpit simulator for crewstation research. The implementation on AIDS will permit investigation of the operation of VRAS in moderately complex flight scenarios.⁸

The VRAS or a similar system may be tested in an airborne platform within the next year. Voice understanding systems have already been tested under conditions simulating airborne noise, G-forces, and vibration using a centrifuge facility.⁹ The NAVAIRDEVCECEN has also taken an Interstate voice recognizer aboard a P-3C aircraft and developed voice recognition patterns during flight.

Outlook

The analytic methods and algorithms developed by the NAVAIRDEVCECEN for performing tradeoff studies may be applicable in the development of ASR training. Some of these techniques, such as MOAT (Mission Operability Assessment Technique),^{10,11} might be applied to assist instructional designers in evaluating ASR tasks to determine training requirements. Others might help in evaluating various media for training ASR users. Certainly, the emphasis on affordability analysis is appropriate.

The VRAS system itself should prove to be a useful tool for research and evaluation. It can be used to investigate aspects of ASR-user interaction using a somewhat constrained but relatively complex syntax. For example, VRAS could be used in a system with a voice input preprocessor for experiments on syntactical variables.

⁸Streib & Preston, op. cit.

⁹Feuge & Geer, op. cit.

¹⁰Helm, W. R. & Donnell, M. L. Mission Operability Assessment Technique: A methodology of manned system evaluation. Point Mugu, CA: Pacific Missile Test Center, 1979.

¹¹Donnell, M. L. The application of decision analytic techniques to the test and evaluation phase of the acquisition of a major air system: Phase III, TR 78-3-25. McLean, VA: Decisions and Designs, Inc., 1979.

Operational implementation of the ASR technology under study at the NAVAIRDEVCEEN is likely to be a gradual process over the next several years. The research programs are structured in phases: a Technology Development phase is followed by Systems Integration and finally by Technology Demonstration. The ASR programs are entering the Technology Demonstration phase, but research is still needed in many areas, such as user acceptance and the effects of ASR implementation on training requirements. Within a shorter time, perhaps within a year or two, we may see less sophisticated implementations as low-cost new technology developments tempt airborne system designers. The research programs should provide guidance, but there is a danger that too rapid implementation of new technology will present problems that could be avoided by a more judicious pace.

NAVTRAEQUIPCEN ASR RESEARCH

Human factors research on ASR at the NAVTRAEQUIPCEN has concentrated on potential applications of ASR to training. However, just as research at the NAVAIRDEVCEEN has produced some findings with implications for training research and application, research at the NAVTRAEQUIPCEN has produced some findings with implications for human engineering research and application.

Air Controller Training

A major portion of the NAVTRAEQUIPCEN ASR research effort has involved training for Precision Approach Radar (PAR) controllers. A computer-controlled adaptive laboratory demonstration trainer showed the feasibility of using ASR for PAR controller training.¹² Subsequently, a prototype Ground-Controlled Approach Controller Training System (GCA-CTS), employing an isolated-word recognition speech preprocessor, was developed under contract to NAVTRAEQUIPCEN and evaluated at the Air Traffic Control Schools, Naval Air Technical Training Center,

¹²Breaux, R. Laboratory demonstration of computer speech recognition in training. Proceedings: 10th NTEC/Industry Conference, Technical Report NAVTRAEQUIPCEN IH-294. Orlando, FL: Naval Training Equipment Center, 1977.

NAS Memphis.¹³ Another trainer, for Air Intercept Controllers, is presently under development and scheduled for evaluation at Fleet Combat Training Center, Pacific, San Diego in early FY 81. It incorporates a more advanced voice processing system, the Nippon Electric DP-100 connected speech processor.¹⁴

GCA-CTS as an ASR System

The GCA-CTS is fully described elsewhere.¹⁵ It provides a good example of a complex, computer-controlled adaptive training system with interaction between operator and machine in the voice mode. Its availability has allowed extensive hands-on experience with the operating characteristics of state-of-the-art isolated word ASR systems.

As part of the present study of the human factors involved in such man-machine voice interactions, one of the authors completed the GCA-CTS Precision Approach Radar controller curriculum. This curriculum was designed to teach student air traffic controllers the procedures and radio terminology used in controlling PAR approaches and landings. The student speaks as if to the pilot and pattern controller, and the GCA-CTS ASR system monitors and evaluates his performance, while providing voice and CRT displays simulating behaviors of the pilot, aircraft, pattern controller, and tower controller. The system

¹³McCauley, M. E. & Semple, C. A. Precision Approach Radar Training System (PARTS) training effectiveness evaluation, Preliminary Final Report NAVTRAEQUIPCEN 79-C-0042-1, Westlake Village, CA: Canyon Research Group, Inc., 1980.

¹⁴Grady, M. W., Hicklin, M. B., & Porter, J. E. AST in the 80's: New systems, new payoffs. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

¹⁵Hicklin, M., Barber, G., Bollenbacher, J., Grady, M., Harry, D., Meyn, C., & Slemon, G. Ground Controlled Approach Controller Training System Final Technical Report. Technical Report NAVTRAEQUIPCEN 77-C-0162-6. Orlando, FL: Naval Training Equipment Center, 1980.

illustrates the potential for "instructorless" training of largely verbal skills.¹⁶

Outlook

The usefulness of the GCA-CTS extends beyond its demonstrated capabilities in the training of air traffic controllers in PAR approach procedures. It can be used as a test system for study of a variety of potential changes in voice recognition and/or training hardware and software. It is instructive to study GCA-CTS as an operating ASR system to understand the human factors at work in its interaction with the user.

Hands-on experience interacting with the GCA-CTS system in systematically selected parts of the PAR controller curriculum could be invaluable to instructional systems designers charged with developing training for ASR system users. It would help them understand the task of interacting with a computer through the voice medium, and thereby provide insight into the selection of appropriate media for training tasks that involve voice technology.

Technology transfer may be facilitated by tapping the knowledge of NAVTRAEQUIPCEN personnel and their contractors who have had experience with the GCA-CTS. This report is intended as a first step toward achieving such transfer. It will describe the lessons learned from hands-on experience with ASR technology and indicate ways in which ISD personnel can share the benefits of that experience as they develop training for and with ASR systems.

¹⁶Breaux, R. Laboratory demonstration of computer speech recognition in training. In R. Breaux, M. Curran, & E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

SECTION III

HUMAN FACTORS CONSIDERATIONS IN ASR TRAINING

Alice opened the door and found that it led into a small passage, not much larger than a rat-hole: she knelt down and looked along the passage into the loveliest garden you ever saw. How she longed to get out of that dark hall, and wander about among those beds of bright flowers and those cool fountains, but she could not even get her head through the doorway; "and even if my head would go through," thought poor Alice, "it would be of very little use without my shoulders. Oh, how I wish I could shut up like a telescope! I think I could, if I only knew how to begin."

17

The intent of introducing Automated Speech Recognition systems into aircraft cockpits will be to reduce aircrew workload and facilitate task performance. The accomplishment of these goals is not straightforward and is likely to be much more complex than is evident from casual reflection.¹⁸ A significant factor in achieving success in the implementation of airborne ASR will be aircrew training, because the full benefits of ASR can accrue only if personnel learn how best to utilize ASR systems.

The introduction of airborne ASR will require adaptation of old behaviors and the learning of new ones by systems operators. The user of airborne ASR technology may find himself in a position somewhat analogous to that of Alice peering through the small passage behind that little door. The full benefits of ASR may be thought of as analogous to the wonders displayed in the beautiful garden beyond the passage. Just as Alice longed to know how to begin to traverse that passageway and wander among the flowers and fountains, so the ASR user is faced with the problem of gaining access to the full benefits of ASR. The human factors peculiar to ASR can act to restrict the user's access, as illustrated in Figure 3. A well-designed training program can provide the "magic" to allow the user access to the garden of benefits.

The requirement for adaptations and new behaviors by the ASR user introduces human factors considerations for ASR training which, for purposes of

¹⁷Carroll, L. Alice's Adventures in Wonderland. New York, NY: Grossett and Dunlap, undated.

¹⁸Lane & Harris, op.cit.

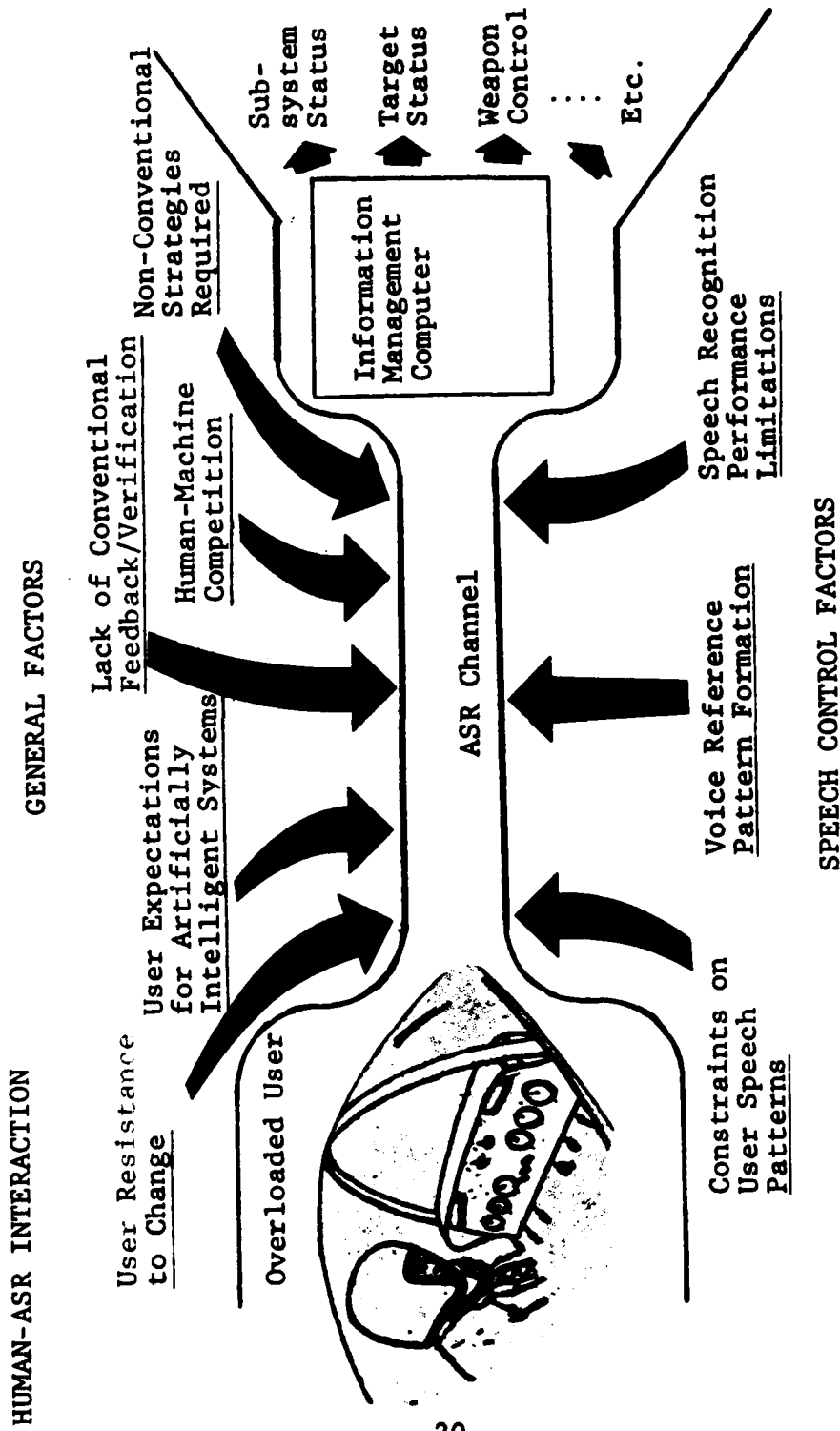


Figure 3. Human factors restrictions on the Automated Speech Recognition Channel.

the present discussion, are analyzed on two levels. The first level concerns the partial shift from the use of manual and visual channels to the use of auditory and voice channels for information exchange between human aircrew members and aircraft systems. It will require significant changes in human information processing techniques and strategies, which will be discussed in detail below.

The second level of human factors analysis concerns changes required in speech patterns and related behaviors involved with use of ASR systems. Here it is necessary to consider the ways in which ASR systems constrain the user's speech, requiring particular speech behaviors which are different from those used in everyday discourse with other people.

The focus of the present discussion will be on human factors associated with introduction of airborne ASR systems. However, those factors are not limited to airborne systems. For example, many of the considerations are likely to be applicable to training systems that use ASR technology, and to a range of other voice-interactive systems.

HUMAN-ASR INTERACTIONS: GENERAL ISSUES

Airborne applications of voice interactive systems will be characterized by a shift from the use of manual and visual channels to the use of auditory and voice channels for information exchange between the human operator and aircraft systems. This shift should assist the aircrew by easing their workload, but if not skillfully managed it could result in an additional burden. The effective use of airborne ASR systems will require careful analyses of the operators' jobs and tailoring of the systems' designs to those jobs.¹⁹ In addition it is likely to entail significant reorientation of user training to teach ASR users new information processing techniques and strategies for use in the ASR-equipped cockpit.

User Resistance to Change

One of the more difficult problems to deal with is likely to be resistance by experienced operators to changes induced by the introduction of ASR technology. To the degree that ASR technology replaces

¹⁹Ibid.

conventional input channels, experienced operators will not be able to interact with their systems in the familiar ways they learned in original training. Thus, a highly organized sequence of behavior learned as an operator task may be interrupted by a requirement to use the novel ASR input mode. The emotional consequences that can follow the interruption of an organized sequence of behavior are well known: interruption may lead to expressions of fear, anger, surprise, or other emotions, any of which can produce further disruption of the organized sequence.²⁰ To the extent that an operator's task is disrupted and the achievement of a mission goal is perceived as thwarted by ASR, the emotions aroused in experienced operators by the introduction of ASR are likely to be negative.

An animal that has learned a simple response, such as running down an alleyway to obtain food in a goal box, will show negative emotional behaviors when the food is no longer forthcoming. If the animal receives repeated exposure to stimuli associated with an empty goal box that formerly contained food, those "empty goal" stimuli may come to have aversive properties and will be avoided.²¹ It is possible that, in a similar way, the operator's formerly friendly cockpit could be perceived as somewhat aversive when ASR is introduced, just because some well learned habits are no longer effective in achieving mission goals. Simply stated, operators may resist ASR just because it is different.

It might be possible to overcome some of this resistance by providing conventional inputs as backup for ASR. Unfortunately, the presence of the conventional backups diminishes the likelihood that the full benefits of ASR will be achieved. The reason for that is a natural tendency for operators to revert to use of the familiar, highly organized and trained behavior if it is available.²²

The solution to the problem of user resistance to change is to provide an effective substitute for the behavior sequences that are no longer available, i.e. an alternative way to complete the tasks.

²⁰Mandler, G. Mind and emotion. New York, NY: John Wiley & Sons, Inc., 1975.

²¹Wagner, A. R. Conditioned frustration as a learnable drive. Journal of Experimental Psychology, 1963, 66, 142-148.

²²Mandler, op. cit.

Thorough training of an alternative response can reduce the likelihood of reverting to a formerly learned behavior that is no longer appropriate.²³ It must be emphasized, however, that mere replacement of conventional input systems with ASR will not suffice. Training will be the critical element in making the alternative behavior, i.e. successful voice interaction, available and preferred by the user.

User Expectations for Artificially Intelligent Systems

The second difficult problem is teaching users the reality of dealing with limited intelligence machines. A computer is a machine that can follow limited instructions. In performing this considerable feat, this artificially intelligent system remains nonetheless a limited machine. For the uninitiated, however, computers have always held a certain aura of mystery. It matters little that the achievements of computers derive only from the ingenuity of their human designers and programmers.

The aura surrounding artificially intelligent systems probably stems from occasions when machines depart from their machine-like predictability to mimic animate or even human functions. It is a common observation that people are intrigued when machines display unpredictability. For example, in describing a pattern of adaptation by a machine model of an animal nervous system, Cofer and Appley comment, "Interestingly enough, the pattern is not predictable..."²⁴

The addition of speech recognition and speech synthesis capabilities to computer systems can only add functions that enhance their status as artificially intelligent. Because the development and programming of voice-interactive systems is a labor-intensive effort that sometimes even involves working around the clock,²⁵ the designers and programmers of these artificially intelligent systems are often painfully aware of the systems' limitations.

²³Leitenberg, H., Rawson, R. A., & Mulick, J. A. Extinction and the reinforcement of alternative behavior. Journal of Comparative and Physiological Psychology, 1975, 88, 640-652.

²⁴Cofer, C. N. & Appley, M. H. Motivation: Theory and research. New York, NY: John Wiley & Sons, Inc., 1964.

²⁵Hicklin, et al., op. cit.

But when the user meets a system that talks and listens, it may seem all too human, and the user may not understand or even be ready to accept that it does not quite measure up to standards set by and for humans.

The attribution of human characteristics to a machine can have consequences that severely impair human-machine interaction. User behaviors that would be appropriate and efficacious in interaction with another human may be quite inappropriate and obstructive in interaction with a machine, even though the machine can mimic some human functions. The most obvious instance of inappropriate and obstructive behavior toward an ASR system is a forceful and tense repetition of a mis-recognized input. This behavior may reflect a natural tendency of response to another human who misunderstands, and that natural tendency may serve well in interaction between humans. Exasperation results when the forceful repetition fails to make the ASR system recognize correctly, making matters worse as discussed later under the heading of Speech Recognition Performance.

Training will be required to help the ASR user to suppress inappropriate human-oriented response tendencies and to strengthen appropriate machine-oriented response tendencies. At the same time, as will be shown in subsequent discussions, the training must include emphasis on making the most of ASR features that offer capabilities transcending the limits usually attributed to machines.

Feedback/Verification of Speech Input

Perhaps the most curious human factors problem is the absence in these systems of many conventional sources of feedback and verification of control inputs. This problem can reduce the operator's certainty of the status of aircraft systems, until a transition is made from more conventional response styles to responses of a more cognitive nature.

For example, when an operator throws a switch or lever to lower the aircraft landing gear, there are several sources of feedback on the input, besides an indicator. When the gear is lowered, there may be a perceptible change in the handling characteristics of the aircraft. Visually, the observed position of the control provides confirmation of the input, and kinesthetically, muscle and joint position cues signal

the accomplishment of the input. If the operator uses a VRAS-type system to "Change aircraft landing gear status to 'down,'" certain visual and kinesthetic sources of feedback will not be present. A change in the aircraft handling feel may still be present, but if the operator thinks he has lowered the landing gear and the ASR system has instead understood "Change aircraft speed brakes status to 'on,'" what will happen? Given the evidence that the perception of an ambiguous stimulus situation is highly susceptible to the influence of explicit set,²⁶ can we be sure how the feel of speed brakes will be perceived by an operator who is set to feel the effects of a lowered landing gear?

The development of a cognitive response style by the operator will be facilitated if the designers of airborne ASR systems are responsive to the operator's need for verification of input. In certain circumstances, such as weapons launch, action on a command demands prior confirmation. The logic of the VRAS system can be configured to require confirmation before acting.²⁷ Acceptance of airborne ASR systems may depend in part on showing the operator in training how the systems provide full access to needed verification, how any limits on verification are justified, and how command confirmation requirements enhance the effectiveness of the systems. The conventional feedback mechanisms such as lights, alarms, and switch positions are replaced with the use of speech communication.

Another more subtle feedback problem may be engendered by any user tendency to attribute human characteristics to ASR systems. The "naturalness" of the voice mode of interaction tends to encourage a communications mode like that used between humans. But many of the natural feedback loops present in human communication are not present in interaction between a human and an ASR system. For example, no currently available ASR system can use maintained eye contact to indicate attention to the speaker. Nor can it "look puzzled" to indicate to the speaker less than full understanding. The VRAS system may process as far as possible into a partially understood statement, and then request further input.²⁸ However, the potential is present in all current ASR systems for

²⁶Dember, W. N. The psychology of perception. New York, NY: Holt, Rinehart and Winston, 1966.

²⁷Lane & Harris, op. cit.

²⁸Ibid.

mis-recognition and incorrect action without providing a person with cues that might indicate trouble and prepare him for corrective action before it is too late. Such cues would be present in speaking to a responsive human listener in proximity. Lacking them, the ASR system may be perceived as unfriendly and threatening, just as an inscrutable and stony-faced human listener would.

Human-Machine Competition

The final problem for purposes of this discussion is evident to some extent in the GCA-CTS. The GCA-CTS as configured for the evaluation²⁹ was characterized by a lack of flexibility in sequencing of training activities by the student. When working through the GCA-CTS curriculum, at least one of the authors found that this characteristic detracted from the acceptability of the system. McCauley and Semple³⁰ used the phrase "locus of control" to refer to the degree of a student's ability to decide upon his own course of training activities, and described the GCA-CTS as a system that left the student uncomfortably passive and controlled by the preprogrammed syllabus.

An airborne ASR system will likely not be set up to control an operator to the extent that the GCA-CTS controls a student in presenting lengthy instructional sequences. However, it may have some acceptability problems if it is perceived by the operator as limiting his control of the situation. The foregoing discussions have illustrated the potential for ASR systems to produce operator resentment just because they represent a new mode of input, and for operators to become exasperated when the human-like machine falls short of full human capabilities. We should not be surprised to find an operator reluctant to relinquish any part of control of the cockpit to a system that arouses such reactions.

²⁹McCauley & Semple, op. cit.

³⁰Ibid.

Non-conventional User Strategies/Techniques Required

There is a solution to the problems just discussed. It is the contention of this report that the right training can furnish the attitude and skills needed by an operator to climb aboard an aircraft with an ASR system and take advantage of its benefits without suffering from the peculiar human factors that such systems share. Training must provide strategies and techniques that the operator can use to realize the full potential of airborne ASR technology.

For ASR systems, the strategies and techniques used by operators will often require novel or somewhat unconventional approaches to task performance. For example, the operator may have to learn to suppress response generalization, that is, the circumstance in which a behavior learned for a situation is prevented and a similar behavior is substituted.³¹ It is normally useful, providing a successful alternative for achieving an otherwise blocked goal. For instance, if on one occasion an operator finds that the normal one-handed pressure on a lever fails to operate it, using both hands and putting more weight on it might succeed in moving it. If a voice input to an ASR system fails to have the desired effect, however, using a different expression or varying the forcefulness of the response decreases the likelihood of success, as will be discussed subsequently under the heading of Speech Recognition Performance.

The use of airborne ASR may require a radical shift in cue dependence. Control function has typically been coded by physical characteristics of the control such as shape, texture, or other features, by location, by label, or by the way the control operates.³² If a system such as VRAS were to become the primary control input for a substantial number of aircraft subsystems, reliance on tactile cues or other

³¹Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York, NY: John Wiley & Sons, Inc., 1951.

³²McCormick, E. J. Human factors in engineering and design. New York, NY: McGraw-Hill Book Company, 1976.

control characteristics for identification would be impossible. When access to controls is through unique verbal commands, special recall techniques may be required to keep track of them. Imagery and mnemonics are being recommended for training Morse Code, Signal Flags, Orders to Sentries, and other technical materials.³³ Perhaps these techniques will find use in interaction with airborne ASR systems.

Lane and Harris³⁴ summarize several studies showing that, especially when an operator works under a high information rate, using voice for display and control functions will increase the performance payoff in weapon systems. With a VRAS-type system, for example, an operator could use a single verbal request for a type of information on all aircraft fuel tanks. To "request" the same information without the ASR system might require a visual scan to check each of several displays, or calling up and scanning information on a multi-purpose display. An airborne ASR system can give the operator flexibility to group requests and commands in ways not possible with conventional aircraft controls and displays. It remains to be seen whether research will determine that certain patterns of information exchange using airborne ASR systems will be most advantageous and should be used by all operators in a given situation, or whether greatest advantage will be conferred by leaving the options open for each crew to select its own preferred pattern. If crews are permitted to exploit the flexibility of ASR systems in their own ways, then training may have to focus less on strict adherence to fixed procedures and more on encouraging continued seeking of novel ways to increase efficiency.

HUMAN-ASR INTERACTION: SPEECH CONTROL FACTORS

Constraints on User Speech Patterns

Although Automated Speech Recognition systems have the potential for great facilitation of communications between human operators and complex machines, it is unlikely to be possible in the near

³³Braby, R., Kincaid, J. P., & Aagard, J. A. Use of mnemonics in training materials: A guide for technical writers. TAEG Report No. 60. Orlando, FL: Training Analysis and Evaluation Group, 1978.

³⁴Lane & Harris, op. cit.

future to speak to a computer just as one would to another person. Current speech recognition systems place some constraints on the operator's speech input, and thus require that the operator be trained to speak in a particular way when using the system.

Human speech may appear at first to be a relatively "natural" free-running behavior, which might be difficult to change or stylize for the purpose of being understood by a computer. In fact, however, all of us continually adapt our speech to the characteristics of those to whom we speak, with very little difficulty. We alter our vocabulary, sentence complexity, rate of speaking, and intonation patterns, speaking one way to an infant, another way to an adult, and still another way to our dog. We slow and simplify our speech for a listener who is hard of hearing, or one who does not know our language well. We use highly technical vocabulary to impress our professional peers, and less complex words to explain our work to a layman. Thus it should be neither unreasonable nor particularly difficult to be asked to adopt a particular style of speech when talking to a computer.

The adaptation of speech to the listener may be conceptualized as the use of an implicit model of the listener's speech understanding capabilities. The characteristics of the model may be based on knowledge of the listener's capabilities, on assumptions about those capabilities, on a population stereotype, or other factors. A speaker tailors his or her speech to fit the model.

For ASR systems, the question becomes one of what characteristics a speaker attributes to the ASR listener. From the viewpoint of the instructional designer, the question must become one of what characteristics operators should be trained to attribute to ASR systems. For the near term, the following factors and constraints will need to be considered by designers of training systems for ASR operators.

1. Stylization. The stylization constraints imposed by an Automated Speech Recognition system are probably the greatest challenge for the developers of ASR operator training. These are the most subtle speech requirements, those which are least obvious to the speaker. As mentioned before, a speaker talking to another person is able to adapt his or her style of speech to the listener's capabilities. In learning to

do this, a speaker relies heavily on cues from the listener that provide feedback on how well utterances are being understood. As discussed under the heading of HUMAN-ASR INTERACTION: GENERAL ISSUES, in the case of ASR many of these cues are not available, and the speaker starts out either with no model of the listener (computer) or one based on preconceived impressions that may be invalid. Thus, ASR training will need to establish for the operator trainee a valid, realistic model of ASR capabilities.

Training for ASR systems may also have to provide instruction in attending to cues which are more subtle than those a speaker uses in adapting to a human listener. Machine understanding of speech can be critically dependent on characteristics of speech to which a speaker normally does not attend and which are not normally thought of as important, such as rate of speaking, or placement of pauses. The ASR user must attend to these characteristics of his or her own speech and use them as feedback cues in order to learn the stylization requirements of an ASR system. To assure sufficient emphasis in training, the instructional designer must have a thorough understanding of the difficulty of teaching speakers to control these characteristics of their speech.

The most obvious example of a stylization constraint is the requirement to pause slightly between words (or phrases which are processed as words) when speaking to an isolated word recognizer. As an extreme example, on the GCA-CTS the operator must learn to say, "Turn right heading (pause) one (pause) five (pause) zero." He must also take care not to insert extra pauses in phrases which are handled as single words by the system: "Turn right (pause) heading..." will not be understood by GCA-CTS. The Japanese have introduced a limited connected speech recognition system (five-word string, maximum), which provides the flexibility to pause or not pause within a group of five words. However, informal evaluation by one of the authors indicates possible confusions by even that system when multi-syllable words are spoken in the same string or utterance with digits.

Stylization also means consistency. The speaker must not vary inflection or volume excessively, because computers find human speech somewhat garbled anyway, and this just makes it worse.

Automated speech recognition system designers may be expected to reduce stylization requirements to a

minimum, but training designers must be prepared to cope effectively with some stylization constraints which cannot be avoided, especially in near-term systems. Conventional training techniques may have to be supplemented by innovative approaches for successful accomplishment of training goals for ASR systems.

2. Vocabulary Constraints. Most current off-the-shelf ASR systems (voice processor, computer memory, and software) which have potential for airborne application can handle 100- to 300-word vocabularies (although expanded systems handling up to 900 or 1,000 words are commercially available).³⁵ The operator trainee must be taught to speak only the words which have had their meaning defined to the system when he speaks to it. If the airborne systems designers have used complete task analysis data when selecting the vocabulary, this should not be difficult; the vocabulary should be adequate for the task requirements. Occasionally, it may be necessary to change some formerly standard terminology in order to avoid confusion among similar-sounding words, such as "for" and "four" or "to" and "two",³⁶ but in most cases it should be possible to retain standard terminology. Thus, vocabulary constraints will not be particularly troublesome, although they will require attention and practice in training.

3. Syntactical Constraints. The syntactical systems or "grammars" incorporated in near-term airborne ASR systems are likely to be much simpler and less flexible than standard English syntax. That is, they will strictly limit the ways in which words can be combined into sentences to be understood by the system. Again, in well-designed systems, such as ones similar to VRAS, the syntax will be as natural as possible, and will incorporate some flexibility, in keeping with task requirements. A VRAS operator, for example, receives an appropriate response, whether he

³⁵Lea, W. A. & Shoup, I. E. Review of the ARPA SUR project and survey of current technology in speech understanding. Los Angeles, CA: Speech Communications Research Laboratory, 1979.

³⁶Stokes, J. M., and Dow, L. Vocabulary Development for the Voice Recognition and Synthesis (VRAS) System. Technical Report 1400.05-A. Willow Grove, PA: Analytics, 1980.

says, "Arm guns" or, "Change guns to armed."³⁷ However, to give the operator that flexibility requires prior establishment of a VRAS vocabulary and syntax allowing those alternative utterances.³⁸ The ultimate utility of the VRAS system is thus dependent on the accuracy of the analysis that serves as the basis for the vocabulary and syntax.

The syntax appropriate to an airborne ASR system will have to be taught to operator trainees, who will need at least some practice to become accustomed to it. If the syntax is very "unnatural", more instructional and practice time will be needed than if it is a more easily adopted grammar. Therefore the success of VRAS is also dependent upon the implementation of a well-conceived training program.

The three types of constraint considered above all are elements of the problem of "habitability" of ASR speech requirements. This problem stems from the general requirement that the ASR operator speak in a particular way to the speech recognition system. Although we know that it is possible to learn to stylize our speech in particular ways for particular listeners, it is also intuitively clear that some constraints will be more easily learned and adhered to than others. Few studies have been done, however, to determine what particular kinds of constraints are most or least habitable. This problem is certainly one which could be resolved with additional research effort, as recommended by others.³⁹

The GCA-CTS and VRAS/AIDS are systems that would serve well as vehicles for research on habitability. The systems have features that allow the manipulation of vocabulary, syntax, and stylization variables, providing training researchers an opportunity to economically and efficiently conduct such research.

Voice Reference Pattern Formation

Most of the presently available ASR systems are "speaker-dependent", that is, they require that each operator "train" the system by providing examples of that speaker's pronunciation of the words to be understood. However, there is now commercially available at least one speaker-independent telephone

³⁷Ibid.

³⁸Ibid.

³⁹Lea, op. cit.

query system,⁴⁰ and the very first airborne applications of Automated Speech Technology may utilize a similar small-vocabulary, speaker-independent ASR approach. Systems such as VRAS and GCA-CTS, because of their vocabularies of 100-300 words, must still rely on speaker-dependent devices and software.

The pattern registration process in speaker-dependent systems generates voice reference patterns in the computer memory which serve as templates to which the recognizer system compares future utterances. A complete explanation of the process is provided by Grady and Hicklin.⁴¹ Briefly, each reference pattern is a composite of the several pronunciations of a given word or phrase entered by the operator or trainee. The formation of reference patterns is extremely important to recognition accuracy, since word or phrase recognition occurs when an utterance is judged by the computer to match one reference pattern better than any other.

It is to be expected, then, that if a word or phrase is spoken in a particular way during reference pattern formation, and then spoken differently later, it may not be recognized correctly. The subtlety of differences which can interfere with recognition becomes clear only after one has attempted to use an ASR device. Differences which are not at all apparent to the speaker may result in non-recognition or mis-recognition of speech, leading to considerable frustration.

For purposes of discussion, it is useful to consider two major sources of variability over time among utterances of the same word or phrase by a single speaker. These are 1) physical context, and 2) psychological context.

⁴⁰Moshier, S. L., Osborn, R. R., Baker, J. M., & Baker, J. K. Dialog Systems automatic speech recognition capabilities present and future. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

⁴¹Grady, M. W. & Hicklin, M. Use of computer speech understanding in training: A demonstration training system for the Ground Controlled Approach Controller. Technical Report NAVTRAEQUIPCEN 74-C-0048-1. Orlando, FL: Naval Training Equipment Center, 1976.

Physical Context. Physical context has long been recognized as a source of variability in speech, and considerable research has been done on the effects of noise, vibration, G-forces, and oxygen mask use on voice quality or recognition accuracy. Summaries of this research appear elsewhere,^{42,43,44} and will not be repeated here. Generally, it is found that ASR systems will perform adequately if voice reference patterns are established under physical conditions very similar to those which will be encountered in actual operation. For example, if registration of voice reference patterns occurs in noise, the system will recognize well in noise, but if voice recognition patterns are trained in a quiet setting, noise during operation may cause reduced recognition accuracy.⁴⁵

Psychological Context. The effects of psychological context have not been studied extensively, but have been noted informally by one of the authors and by many other ASR researchers during ASR workshop discussions.^{46,47} Perhaps the most widespread observations are 1) that words trained individually may not be recognized when later embedded in longer utterances, and 2) that a speaker is often mis-recognized when speaking in a stressful situation if his voice patterns have been entered in a non-stressful setting. Such mis-recognition may be self-perpetuating, since it induces additional stress,

⁴²Feuge and Geer, op. cit.

⁴³Lea, op. cit.

⁴⁴Coler, C. R. Automated speech recognition and man-computer interaction research at NASA Ames Research Center. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, Texas, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

⁴⁵Drennan, T. G. Voice technology in attack/fighter aircraft. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

⁴⁶Breaux, R., Curran, M., & Huff, E. (Eds.) Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

⁴⁷Harris, S. (Ed.) Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

which leads to further mis-recognition. This problem will be discussed further under the heading of Speech Recognition Performance. If speech recognition is to work well in a variety of psychological contexts, it is probably necessary to perform voice reference pattern formation under conditions that effectively simulate the range of operational situations to be encountered. In some cases, the actual operational setting may be the most practical site for collection of voice reference patterns. However, it may be possible to obtain speech samples which are sufficiently typical of the trainee's normal voicing by collecting them during the practice of correct terminology in ASR training, as was done for parts of the GCA-CTS vocabulary.⁴⁸

The need for voice reference pattern collection is seen by some researchers and planners as an obstacle to the adoption of ASR systems. If, in fact, every user had to provide, say, ten repetitions of every word in his system's vocabulary each time he went to use a new station, it certainly would be an obstacle. Although the Japanese again have introduced a system which nearly eliminates the need for repetitions, an alternative that may be acceptable is to have each user create a tape cassette or diskette record of training utterances which can be quickly entered in any station to be used. Alternatively, for stations where a limited number of users are encountered (e.g., all planes of a particular squadron), such records for all authorized users could be stored in the system computer, and accessed by a simple user code for each operator as he "signed on" to the system. This would be compatible with the "Crew-Adaptive Cockpit" concept.⁴⁹ Thus the need for voice reference pattern collection, while it may be an inconvenience, need not prevent effective use of airborne ASR systems. As mentioned earlier, the first airborne systems may even be speaker-independent. Certainly, the manner in which voice pattern registration is handled in a particular system will have implications for the training of that system's users. Training design personnel will have to understand the user requirements involved in various approaches to voice reference pattern formation, and design training appropriate to such requirements.

⁴⁸Hicklin, et al., op. cit.

⁴⁹Reising, J. The crew-adaptive cockpit: Firefox, here we come. Proceedings of the Third Annual Conference on Digital Avionics Systems, Dallas, TX, 1979.

Speech Recognition Performance

All currently available Automated Speech Recognition systems have performance limitations which render them less efficient and less adaptable than a human listener. Although some limitations are not easily surmounted, others stem from conditions which can be controlled to minimize their detrimental effects on recognition.

Perhaps the most important of these controllable factors is the design and execution of the voice reference pattern formation procedure. For optimum recognition accuracy, the reference inputs must match the later operational inputs as closely as possible. If operational inputs will be variable, reference inputs should vary over the same range, as previously discussed in the section on voice reference pattern formation. The problem of mismatch between voice reference patterns and later inputs can occur even in "speaker-independent" systems, if the reference patterns which are programmed a priori do not represent the voice types, speech patterns, and noise conditions which will be encountered in operation. Although the software design strategy determines the range of variation permitted for each reference pattern, the key to successful speech recognition performance may be to reproduce for voice reference pattern formation the physical and psychological context under which recognition will have to occur.

Another factor known to affect recognition accuracy is variability among individual users in their ability to "talk to a machine". Some users are consistently well understood by ASR devices, while others have persistent problems, probably because their speech is more variable.⁵⁰

⁵⁰Doddington, G.R. Speech systems research at Texas Instruments. In R. Breaux, M. Curran, and E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffett Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

Perhaps this could be dealt with as a training problem: one could seek to identify and "train out" those speech characteristics which interfere with good ASR accuracy. It is not presently known whether this approach is feasible. An unacceptable alternative would be to select only the well-understood candidates to be ASR users. This is, of course, not likely to be practical when airborne systems are implemented model-wide, since flight school personnel might reasonably question the validity of selecting their students on the basis of their speech quality.

There is also informal evidence suggesting that expectancy plays a role in individual differences in ASR recognition accuracy. Those potential users who expect to be understood generally are relatively well understood by ASR systems, while those who expect the worst from a system generally get it. This problem could be attacked through special training to improve the performance of those who are poorly understood, which might or might not be effective. Alternatively, a public relations effort, perhaps as a part of training, might help increase expectancies for successful recognition. Great care would have to be taken to avoid overselling, or creating unrealistically high expectations, in this case. The absolute and relative effectiveness of these alternative approaches, or of others which might be devised, remains a subject for research.

If we acknowledge that currently available ASR systems, and systems likely to be fielded in the near future, do not always recognize speech with high accuracy, it becomes necessary to assess the effects of non-recognition or mis-recognition on the performance of man-machine systems. We shall concentrate here on the effects on the user and on interactions with the system. The problem of detection of, and aircraft system response to, improperly understood commands must be dealt with by ASR systems designers, and requires exacting human factors analyses.

The first and most obvious effect on the user of recognition failure is that he or she becomes frustrated. Often, a speaker who is not understood reacts by speaking louder and perhaps more quickly, especially if there is time pressure, as in several of the GCA-CTS tasks. The speaker's voice quality may reflect stress or annoyance. Naturally, to the extent that all of these characteristics are not represented in the voice reference patterns, they increase the

probability that the next utterance will be mis-recognized. This sets up a "positive feedback loop" in the cybernetic sense,⁵¹ where mis-recognition leads to speech changes which lead in turn to further mis-recognition. To break this loop, it will be necessary to provide the operator with instruction in responding to mis-recognitions. The next section will present some guidelines for development of training to include such instruction, and also training designed to reduce the user's emotional reaction to recognition failures, if possible, since this reaction appears to underlie the "positive feedback loop" or vicious cycle behavior.

Besides the immediate effect of mis-recognition, there is a more generalized effect on the user's attitude toward the ASR system. Repeated recognition failures may lead the user to lose confidence in the system's competence, and to react negatively to the ASR situation. As we have mentioned before, this negative attitude may lead to poor recognition performance, starting another vicious cycle.

The most disturbing situation for the ASR user, in the opinion of the present authors, is one where the ASR device fails to recognize correctly, but the operator is not given enough information feedback to know that it has done so. Such poverty of feedback is particularly troublesome in a trainer such as the GCA-CTS, because of the trainee's inexperience with the task being trained. Unlike an expert user, the trainee is likely to have difficulty discriminating between incorrect recognition by the ASR system and incorrect behavior on his part.

Although users of airborne ASR systems will usually be experienced, or at least familiar with their tasks through training, feedback on recognition accuracy is important. For example, if a pilot of a two-engine aircraft suspects a problem with one engine, he might ask a VRAS-type system to "Report aircraft engine temperature one." If a mis-recognition occurs, the VRAS system may return "Engine temperature two is ...", giving the pilot enough information to detect the mis-recognition of engine number. If it simply returned a temperature reading, the pilot could not detect a mis-recognition, and might take dangerous actions based on the incorrect information.

⁵¹Van Cott, H. P. and Kincade, R. G. (Eds.), Human engineering guide to equipment design. Washington, D.C.: American Institutes for Research, 1972.

A well-designed ASR system notifies the operator that it has not understood, or displays what was understood to have been said, thus giving the speaker a chance to detect and cope with a mis-recognition problem. The speaker can repeat the utterance, or retrain the system if necessary. But a system which merely fails to respond appropriately to an utterance leaves the operator not knowing what has gone wrong, nor what can be done to set it right. This situation is extremely frustrating, and has a strong negative influence on the operator's attitude.

Problems of this sort can be overcome by several strategies. One is good ASR system design, mentioned above, which gives the speaker enough information to allow him to adapt his behavior smoothly to the system's requirements. At present, approaches to development of training strategies for ASR can best be learned by instructional designers through hands-on experience with speech systems such as GCA-CTS and VRAS. However, some general principles which will facilitate learning from such hands-on experience can be stated and are presented in the next section.

SECTION IV

GUIDELINES FOR THE USE OF ASR IN TRAINING: OVERCOMING HUMAN FACTORS PROBLEMS

The human factors problems raised by the use of ASR in airborne systems have been discussed in Section III, along with some implications for training the operators of ASR systems. This section presents some informal guidelines for the design of training for ASR operators. Section V will consider ways in which these guidelines can be integrated with the ISD process when ASR training systems are designed.

For purposes of discussion, the process of training for and with ASR will be broken down into phases. In an actual training program, these would not necessarily be separate steps in the training process, but they represent three logically distinct functions of a training program: 1) Introduction, 2) Speech Discipline, 3) Principles and Strategies.

INTRODUCTION

The first phase is the introduction and initial presentation of Automated Speech Recognition to a new trainee. The major function of this phase is to familiarize the trainee with the operation of an ASR system, and to develop positive but realistic expectations for successful interaction with the system. It is important in this phase to demonstrate the successful use of an ASR system, showing its benefits and capabilities. Training designers must thoroughly understand the features of the particular ASR systems which trainees will be using, in order to convince their potential users of their value. At the same time, they must be aware of the systems' limitations, and carefully avoid overselling. Establishment of unrealistically high expectations can only lead to later disappointment and loss of confidence in the system.

Another objective of the initial introduction is to begin teaching the trainee how to control the system, and to demonstrate that the operator is in control. It is important to do this early in training, to avoid user suspicion that the system may "take over", controlling or constraining human performance of his tasks. Where the ASR system allows

it, trainees should be shown the flexibility of the system, and the ways in which it can adapt its responses to their needs.

The introductory phase of training also provides the first opportunity to show trainees a model of good ASR speech habits, which they must emulate. Training designers should take advantage of this opportunity, using whatever medium is feasible (e.g., film, videotape, audio tape, live demonstration, etc.) to show examples of the skillful use of ASR. Trainees will almost invariably model their behavior after whatever implicit or explicit examples are provided, as was seen in the GCA-CTS evaluation, where some trainees imitated the synthesized voice.⁵² Thus it is imperative that a good model be provided throughout training, starting from the very beginning. In summary, the introductory phase of training should be used to build the trainees' confidence in the system, dispel their suspicions, and begin to establish the behaviors needed to use the system successfully.

SPEECH DISCIPLINE

The second training phase to be discussed is the speech discipline phase, where trainees learn to speak in a manner that maximizes successful understanding by the ASR system. This phase is likely to be the most difficult for the trainees, and will require skillful use of innovative instructional techniques. The distinguishing features of good machine-recognizable speech behavior still are not well understood, and research in this area could yield findings of great importance to this phase of training.

Assuming that ISD personnel are able to define behavioral objectives for producing good machine-recognizable speech, they will have to provide students with evaluation of their speech behavior in a form (or forms) which students can utilize to modify their speech. Again, a model or example of good speech behavior is the first requirement. It would seem useful to have some means to identify, and explain or display to students, the ways in which their speech differs from the ideal. Displaying to the trainee the word or phrase that was understood to have been said has been used in some approaches, such as the "voice test" mode on GCA-CTS. Unfortunately, this mode on the GCA-CTS provides the trainee only

⁵²McCauley and Semple, op. cit.

information about what utterance was recognized, and little if any helpful information about relationships among utterances. Furthermore, the trainee must infer, by several trials of an utterance, how reliably the system recognizes that utterance.

The potential exists for more creative, sophisticated approaches. Many off-the-shelf ASR systems can display the probability with which a spoken word matches candidate words in its vocabulary. Such information certainly could be used to inform the student about utterances that are hard for the ASR system to distinguish, and indicate where change is needed or where re-registration of voice reference patterns might help. Thus, if a trainee spoke "five", and the ASR system displayed:

Nine - 50 percent
Five - 40 percent
Fire - 10 percent

the trainee would have more usable information with which to modify his pronunciation than if it merely echoed "Nine." An even better approach for some applications would be for the ASR to report periodically to the trainee a list of items which have been having very close probabilities. The trainee might then choose to initiate new voice reference pattern formation.

It may be feasible to give more detailed error feedback, with suggestions for correction, using current ASR technology with new software. Given an appropriate research and programming effort, it is possible to foresee even more exciting potential for using probabilities of match between utterances and voice reference patterns. One interesting possibility for research would be to try training operator speech using the behavior modification principle of shaping, or reinforcement of successive approximations to voice reference patterns. Although the capability to support this type of training of utterances is not currently implemented on any ASR system, the concept of verbal behavior as subject to control according to the basic principles of learning is well over twenty years old.⁵³ However, a research effort on the order of two man-years would likely be needed to determine the behavioral parameters that contribute to recognizability of speech by different ASR systems,

⁵³Skinner, B. F. Verbal behavior. New York, NY: Appleton-Century-Crofts, 1957.

the behavioral factors in speech which produce utterances confusable to current ASR systems, and ways of altering utterances to reduce confusability. A substantial programming effort would be needed to develop a system to use the information produced by such research in an ASR speech trainer.

In an ASR system which provides feedback such as the probabilities with which matches to reference patterns are made, such feedback might serve as differential reinforcement for speech behavior, allowing a speaker to gain control over speech characteristics not usually consciously varied, just as a person can gain control over brain alpha rhythm through biofeedback.⁵⁴ This might work to overcome problems with potentially confusable utterances, or even to help persons who start out with low success in being recognized by ASR systems. It might also be useful to provide such capability in certain operational ASR systems as a means of operator refresher training. This technique would be appropriate for both speaker-dependent and speaker-independent ASR systems.

The provision of accurate, helpful feedback, necessary as it is, is not sufficient to ensure the achievement of successful ASR speech recognition, at least with a speaker-dependent system. An effective technique for the establishment of voice reference patterns is equally important. If voice reference patterns for later operational use are to be recorded during training, the process must be carefully controlled. If this is not done during training, then trainees must be taught the skills to do it later in the absence of the instructor.

In the opinion of the authors, there are some guidelines which should be followed during training and reference pattern formation to ensure that voice reference patterns will provide a basis for good recognition in the operational situation. The first is that the physical and psychological contexts must match, at least in critical dimensions, those to be encountered in operation. Replicating the physical setting should be fairly straightforward, especially if a Flight Trainer is available. Careful attention

⁵⁴Rachlin, H. Behavior and learning. San Francisco, CA: W. H. Freeman and Company, 1976.

must be paid to details such as type and variability of noise, vibration, and G-forces,⁵⁵ but the technology needed for such simulation is available.

The matter of psychological context is more challenging, since there are not enough data presently available to identify the critical psychological contextual variables for voice recognition. Lacking such data, a temporary solution is to replicate the operational context as closely as possible, within the limits of cost and common sense. Utterances to be used for voice reference pattern formation recording should be prompted in the same mode as they will be in operation, so far as is possible, whether it be vocal, printed display, or memory. Words should be embedded in operational utterances, not read individually from a list. To replicate the emotional tone of operational situations may be more difficult, but it should not be impossible. If the ASR will have to understand an operator in stressful situations, for example, then at least some stress should be induced during reference pattern formation. Again, a scenario presented in a Flight Trainer may be sufficient to provide the appropriate context.

The authors realize that training designers may wish to minimize the amount of training time spent in voice reference pattern formation, especially in the formation of reference patterns which are for use only in training. It is indeed desirable that trainees spend as little time as possible in speaking for the sole purpose of registering voice reference patterns, an activity with little training value to the trainee. However, if voice reference pattern registration is well integrated into the training program, it can occupy considerable time, and that time will also be beneficial to the trainee. It is essential that this be done to avoid wasteful use of trainee time, and also to avoid trainee boredom or loss of interest. Fortunately, such integration of reference pattern registration into substantive training exercises also serves the purpose of ensuring the proper context for reference pattern registration.

One final comment on voice reference pattern formation during training is in order. Since the trainee will, in an effective training program, be constantly improving and changing his speech, a good training program will include frequent updating of reference patterns. This process may be performed

⁵⁵Coler, op. cit.

openly, in such a way that the trainee knows that he is updating reference patterns (and giving the trainee a measure of control over the process), or it may be integrated into the training so as to be transparent or unnoticed by the trainee. If the updating is transparent, there must also be a provision whereby the trainee can deliberately test and update the voice reference patterns if poor ASR understanding occurs at any time during training. GCA-CTS has such a provision, although it is somewhat inconvenient to use as it is now programmed, and would benefit from changes allowing the trainee more direct control over the timing and extent of re-registration of voice patterns.

PRINCIPLES AND STRATEGIES

In this final phase of training, the operator trainee, who is now able to "talk to the airplane", will be taught when to talk to it, and how to use the voice system in operation. This training phase will provide instruction in how to utilize the capabilities of the ASR system to perform the job more easily and more effectively than would be possible without ASR. The trainee will learn new ways of obtaining information about the aircraft and its environment, and new ways of entering data or giving commands to the aircraft. In this phase there will be some extension of speech discipline learning, since the trainee will now have to be made comfortable with the vocabulary and syntax limitations of the ASR system.

In designing this phase of training, ISD personnel will need accurate task analysis data, from analyses which have specifically considered how ASR best can be employed in the trainees' task performance. They then will have to design instructional and practice materials which will ensure that trainees learn how to take advantage of the ASR system's capabilities, and how to deal with its limitations.

Whenever possible, trainees should be shown alternative ways to use ASR, and encouraged to practice until each develops a personal style of interaction which works well and is personally "habitable." Of course, if a particular ASR system has little flexibility, training for it must ensure that trainees learn to adhere to stricter constraints on their modes of information transfer. Given an

alternatives-oriented ASR system such as VRAS, however, operators will have a variety of ways to ask for information, a variety of information output modes and formats, and a variety of data entry modes or formats. The trainees must be made aware of the benefits and limitations of each of these, with explicit instructions for situations where one alternative is clearly the most or least suitable. Trainees then should be encouraged to experiment with these alternatives in simulated mission scenarios, until each finds the strategies which provide the best assistance in completing mission objectives.

Experimenting with alternatives in mission scenarios should help convince the trainee that the ASR system will permit maintenance of control over cockpit information handling, a conviction that should contribute to satisfaction with the ASR system. It should help avoid the problem, discussed in section III, of the operator coming to believe that the ASR or computer has "taken over" some tasks, or to perceive that it threatens human control of the mission.

Another important objective of this phase of training is to prepare the trainee to respond intelligently to failures of the ASR system, especially to recognition errors. Again, ISD personnel will need to consider what strategies they wish to teach for a particular ASR system. Thorough analysis of the impact of recognition failure at various points in task performance will be required to determine objectives for this part of training.

Recognition failure is more critical for some situations than for others. For example, during an approach exercise on the GCA-CTS, timing of verbal responses to the system displays and simulated communications is critical. A recognition failure during the approach frequently results in loss of control of the simulated aircraft and considerable trainee frustration. For a less time critical task, repetition of a mis-recognized input may be possible with little change in task success and much less frustration. However, certain principles apply to nearly all ASR systems. Operators must be taught to anticipate some mis-recognition, and to be alert to the cues from the ASR system which indicate that an utterance has not been understood correctly. They must learn to control emotional behavior in the presence of frustration, since such behavior can only aggravate recognition problems. Finally, operators must learn the alternatives available to them when

mis-recognition occurs, and must learn how to choose the best alternative in a given situation. For instance, in a non-critical flight situation, there might be ample time to do a voice test and re-register a word or phrase if necessary. In a critical mission phase, however, the best response might be to repeat the misunderstood word once and then, if that proves unsuccessful, to go immediately to a manual back up system.

These principles are the same that apply in responding to any system malfunction, and may be taught in the same way as other corrective or emergency procedures are taught. As with any such procedures, ample practice under simulated operational conditions should be included in training. If the initial introduction to ASR has been handled skillfully, the operators' expectations of the system will be realistic, and malfunctions of ASR should be no more disturbing than malfunctions in another aircraft subsystem of equal criticality.

SECTION V

IMPACT OF AUTOMATED SPEECH RECOGNITION TECHNOLOGY ON MEDIA SELECTION AND OTHER ISD PROCEDURES

This section discusses for hands-on and academic media selection and other ISD procedures the impact of operational implementation of ASR technology. It also includes a subsection on the use of ASR as a medium for possible application in training for any of several Navy jobs with a substantial speech component.

Automated Speech Recognition technology presents both a challenge and an opportunity for instructional designers. The challenge is to assure that training for airborne and other operational ASR systems is designed to take account of the peculiar human factors of those systems and to prepare operators to cope with those factors. The opportunity is to bring new instructional power and cost savings to various training applications through exploitation of ASR technology to automate some instructional functions usually performed by human instructors or trainees. The challenge will be addressed first.

INSTRUCTIONAL SYSTEMS DEVELOPMENT FOR OPERATIONAL ASR

The procurement of training for airborne ASR is not anticipated to require departures from MIL-T-29053A(TD) dated 14 December 1979. However, the human factors discussed in section III and the training considerations presented in section IV are likely to have a significant impact on the selection of media for accomplishing hands-on training objectives. In the development of training for ASR systems, other ISD front-end analyses preceding media selection will also be affected. For these reasons, those persons responsible for the development of training for operators of ASR systems must themselves have some first-hand experience with ASR technology.

Hands-on Media Selection

Section III of this report mentioned the benefits of hands-on experience with ASR to allow instructional designers to become familiar with the human factors of ASR technology. It goes without saying that front-end

analyses for airborne ASR system training are likely to find hands-on experience essential for training operators of airborne ASR systems. For purposes of the present discussion, it is assumed that implementation of airborne ASR will produce at least some learning objectives which will require hands-on training and practice, with real or simulated ASR equipment.

The definition of "hands-on" training is somewhat unclear for ASR systems. Certainly it implies practice in speaking to an ASR system, so perhaps "voice- and ears-on" would be more descriptive terms. However, in airborne applications, it is likely that ASR systems will usually be integrated with manual data entry/retrieval systems and other conventional controls and displays. Thus it becomes necessary to consider reflecting such integration in training simulation systems, to fulfill objectives which require trainees to learn how to choose and utilize various information exchange channels and formats. For example, it may be anticipated that, as airborne ASR systems begin to be implemented, cockpit procedures trainers, flight simulators, and other trainers will come to include ASR systems.

If an airborne system using ASR is to be simulated with an ASR system different from the actual airborne one, the training development effort must, of course, assure that the ASR system used in the simulator is suitable to the training objectives. It may sometimes be possible to meet training objectives for a costly airborne ASR system through use of a less expensive ASR system if it can provide training experiences of sufficient psychological fidelity to the airborne system.

Currently available ASR devices (voice processor, computer memory, and software) range in cost from a few hundred dollars, for small-vocabulary, low-accuracy recognizers produced primarily for the hobbyist market, to about \$80K for large-vocabulary, high-accuracy systems.⁵⁶ At the lower end of the price range are isolated-word recognizers which have a vocabulary limit of perhaps a few dozen words. The upper end of the range includes the new multi-channel, continuous speech recognizers, which can handle short digit strings and other simple connected utterances.

⁵⁶Lea, op. cit.

Academic Media Selection

For academic media selection, it may be anticipated that in training for ASR systems it will be desirable to exploit media permitting audio recording and playback. Certainly, some ASR knowledge demonstration objectives may be achieved through use of print or other visual media, but demonstrations of voice interaction and evaluative feedback on trainee speech behavior, as discussed in section IV, is likely to require auditory presentation.

At least some speech behavior training is likely to be integrated with a hands-on task trainer, perhaps even a sophisticated automated adaptive trainer such as the GCA-CTS. When it is, the line between academic media and hands-on (voice- and ears-on) media becomes somewhat indistinct. For example, the GCA-CTS contains a voice recording and playback capability that is used in prompting and feedback on trainee speech inputs for voice reference pattern formation. To consider this feature an academic medium seems too limited, yet it is not strictly automated speech recognition or synthesis.

Other ISD Procedures

First-hand experience with ASR technology seems a necessary basis for personnel responsible for phases of training system development other than media selection. For example, section III mentioned the importance of having the training development based on accurate task analysis data from analyses that examine the integration of ASR in task performance. Section VI will recommend basing it on the human engineering task analyses or training task listings for the prime system. Experience with ASR technology will help the instructional designer in applying the task listings (section 3.4 of MIL-T-29053A(TD)) to development of training objectives. Furthermore, unless the instructional designers have personal experience with ASR technology, there is a danger that some important factors in learning about ASR will not be reflected in the development of instructional objectives and their hierarchies (section 3.8 of MIL-T-29053A(TD)).

ASR AS A TRAINING MEDIUM

Although the focus of much of this report has been on training for airborne applications of ASR technology, the advent of ASR and automated speech synthesis presents media that put powerful new tools into the hands of designers of instructional systems for other jobs involving speech. Not only do these media offer advantages for improved standardization and efficiency of training, but they can yield significant cost savings through replacement of certain voice interactive functions normally performed by instructors.⁵⁷

The Air Force is beginning to use voice technology in production simulators for the F-4E and A-7D aircraft.⁵⁸ Sophisticated technology has been developed⁵⁹ and evaluated.⁶⁰ Its transfer represents for instructional designers the opportunity mentioned at the beginning of this section. With creative application, ASR could become a fixture in training programs for Air Controllers, Radar Intercept Officers, Officer of the Deck in ships operations, and other speech-based jobs.

⁵⁷Breaux, 1977, op. cit.

⁵⁸Grady, Hicklin, & Porter, op. cit.

⁵⁹Hicklin, et al., op. cit.

⁶⁰McCauley & Semple, op. cit.

SECTION VI

CONCLUSIONS AND RECOMMENDATIONS

GENERAL REMARKS

Automated Speech Recognition (ASR) technology may soon be airborne and is currently ready for implementation in training applications. ASR presents human factors challenges which must be answered intelligently, but which should not preclude its productive application to a variety of military training situations in the near future. The following recommendations, divided into two sets, suggest solutions or approaches to solutions for some of the human factors challenges identified in section III of this report. The first set includes recommendations relevant to any use of ASR in training, while the second set is applicable principally to training for the operational use of ASR systems. Each set includes some general but basic suggestions for training, followed by more specific prescriptions concerning the content of training. Following each recommendation is a reference to pages in preceding sections of this report where further rationale and discussion on the topic can be found.

ASR IN TRAINING SYSTEMS

The following four recommendations apply to the development of training for jobs with a significant speech base, such as Officer of the Deck in ships operations, Air Controller, and other Naval Flight Officer positions. Training for such jobs could include ASR capability. Currently available technology (e.g. GCA-CTS) in its present configuration or preferably with selective modifications can support the implementation of these recommendations.

1. Hands-on ASR

Instructional system designers, especially those who are charged with the development of training which may employ ASR, should obtain some hands-on experience with ASR technology. This will introduce them to a new medium with potential for cost savings and improved training (see p. 51), and will ensure that these personnel are at least acquainted with the human

factors of working with an ASR device, and that they are able to consider the operator's point of view when making design decisions (see p. 18)

2. Speech Behavior Models

Any training system employing ASR should provide some demonstration of ASR speech behavior samples and their effects on recognition. Examples of correct and highly machine-recognizable speech are especially important for the trainees to emulate. Trainees using ASR systems typically model their speech on the examples available to them. Therefore, these examples should be chosen to illustrate specific factors as effectively as possible (see p. 41).

3. Speech Evaluation and Feedback

Any training system employing ASR should provide an effective and easily accessed means for trainees to evaluate their own speech behavior, and to receive informational feedback on its quality or its intelligibility to the ASR system. This is especially important with speaker-independent systems, which are not very adaptable to individual speakers (see p. 38 and pp. 41-43).

4. Recognition Test and Voice Reference Pattern Update

Any training system employing speaker-dependent ASR should provide a convenient means by which trainees can test voice recognition and update voice reference patterns. This capability will aid in preventing the frustration which arises from incorrect recognition, and will foster trainee perceptions of control over ASR functions (see pp. 44-45).

TRAINING FOR OPERATIONAL USE OF ASR

The following five recommendations apply to the development of training for operators of airborne or other operational ASR systems. They may also be relevant to ASR systems used only for training, but some of them are less critical in that context.

1. Human Factors Training Analysis

Training programs for new users of Automated Speech Recognition systems should be based on front-end analysis data developed with the

participation of professional personnel who have a thorough understanding of the human factors of ASR. This will help ensure that those factors are represented correctly in training and that trainees are taught effective responses to human factors problems engendered by ASR technology in the operational setting (see p. 50).

2. ASR Integrated in Task Performance

Training programs for operational ASR systems should be based on data from human engineering task analyses or training task listings which specifically address the integration of ASR in overall airborne task performance. The intent of this recommendation is to prevent ASR being presented as an add-on "gadget," and to ensure that trainees learn to use ASR in the proper context in task performance (see pp. 27-28 and p. 50).

3. Personal Style in ASR Use

Training programs for operators of airborne ASR devices should include instruction in seeking alternative uses for ASR in task performance, and trainees should be encouraged to develop a personal style to optimize their performance of aircrew tasks. In combination with emerging digital avionics systems, ASR opens many options for crew-aircraft information exchange, and trainees will need time for guided experimentation to find appropriate combinations of options which work best (see pp. 45-46).

4. Voice Reference Pattern Formation Context

Training programs for new users of speaker-dependent ASR systems should provide instruction and practice in voice reference pattern registration. The registration of voice reference patterns which are representative of the operational context is critical to recognition performance, and must be given high priority in training. Training may be able to provide the best physical and psychological context for actual voice reference pattern formation. If reference patterns for operational use are not to be registered during training, then trainees must be thoroughly prepared to perform the registration later on the job (see pp. 34-35 and pp. 43-44).

5. Recognition Failure Experience

Training for operators of operational ASR systems should provide explicit instruction and practice in coping with recognition failures. Trainees should be taught a variety of responses and criteria for choosing the best response to recognition failure in a given operational situation (see pp. 37-38 and pp. 46-47).

A FINAL COMMENT

The authors' hands-on experience with the GCA-CTS and VRAS voice technology systems has provided considerable insight into the human factors problems which are characteristic of ASR systems. This hands-on experience suggests that many of these problems stem from design limitations, and might be susceptible to improvement by appropriate human engineering design changes. Evaluation by others⁶¹ has also suggested design modifications. The question arises of why these design limitations take the form they do, and whether other systems under study or development might show less encumbering limitations.

Chatfield, Marshall, and Gidcumb⁶² presented persuasive arguments for increasing the flow of information between basic researchers and contractors that produce voice technology systems. It is suggested here that productive interchange could be achieved through a workshop or workshops attended by those involved in research, production, and evaluation of voice technology systems. Useful interchange for all participants might be fostered by discussions which have been directed to focus on specific topic areas.

⁶¹McCauley & Semple, op. cit.

⁶²Chatfield, D. C., Marshall, P. H., and Gidcumb, C. F. Instructor model characteristics for automated speech technology (IMCAST). Technical Report NAVTRAEQUIPCEN 79-C-0085-1. Orlando, FL: Naval Training Equipment Center, 1979.

REFERENCES

Braby, R., Kincaid, J. P., & Aagard, J. A. Use of mnemonics in training materials: A guide for technical writers. TAEG Report No. 60. Orlando, FL: Training Analysis and Evaluation Group, 1978.

Breaux, R. Laboratory demonstration of computer speech recognition in training. Proceedings: 10th NTEC/Industry Conference, Technical Report NAVTRAEQUIPCEN IH-294. Orlando, FL: Naval Training Equipment Center, 1977.

Breaux, R. Laboratory demonstration of computer speech recognition in training. In R. Breaux, M. Curran, & E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

Breaux, R., Curran, M., & Huff, E. (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.

Brogden, W. J. Animal studies of learning. In S. S. Stevens (Ed.), Handbook of experimental psychology. New York, NY: John Wiley & Sons, Inc., 1951.

Carroll, L. Alice's Adventures in Wonderland. New York, NY: Grosset and Dunlap, undated.

Chatfield, D. C., Marshall, P. H., and Gidcumb, C. F. Instructor model characteristics for automated speech technology (IMCAST). Technical Report NAVTRAEQUIPCEN 79-C-0085-1. Orlando, FL: Naval Training Equipment Center, 1979.

Cofer, C. N. & Appley, M. H. Motivation: Theory and research. New York, NY: John Wiley & Sons, Inc., 1964.

- Coler, C. R. Automated speech recognition and man-computer interaction research at NASA Ames Research Center. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, Texas, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.
- Curran, M. Voice integrated systems. In R. Breaux, M. Curran, & E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffet Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.
- Dember, W. N. The psychology of perception. New York, NY: Holt, Rinehart and Winston, 1966.
- Doddington, G.R. Speech systems research at Texas Instruments. In R. Breaux, M. Curran, and E. Huff (Eds.), Proceedings: Voice Technology for Interactive Real-time Command/Control Systems Application. NASA Ames Research Center, Moffett Field, CA, 1977. Reprinted by Naval Air Development Center, Warminster, PA, 1978.
- Donnell, M. L. The application of decision analytic techniques to the test and evaluation phase of the acquisition of a major air system: Phase III, TR 78-3-25. McLean, VA: Decisions and Designs, Inc., 1979.
- Drennan, T. G. Voice technology in attack/fighter aircraft. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.
- Feuge, R. L. & Geer, C. W. Integrated applications of automated speech technology final report, ONR-CR213-158-1AF. Arlington, VA: Office of Naval Research, 1978.
- Grady, M. W. & Hicklin, M. Use of computer speech understanding in training: A demonstration training system for the Ground Controlled Approach Controller. Technical Report NAVTRAEQUIPCEN 74-C-0048-1. Orlando, FL: Naval Training Equipment Center, 1976.

Grady, M. W., Hicklin, M. B., & Porter, J. E. AST in the 80's: New systems, new payoffs. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

Harris, S. (Ed.) Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

Helm, W. R. & Donnell, M. L. Mission Operability Assessment Technique: A methodology of manned system evaluation. Point Mugu, CA: Pacific Missile Test Center, 1979.

Hicklin, M., Barber, G., Bollenbacher, J., Grady, M., Harry, D., Meyn, C., & Slemon, G. Ground Controlled Approach Controller Training System Final Technical Report. Technical Report NAVTRAEQUIPCEN 77-C-0162-6. Orlando, FL: Naval Training Equipment Center, 1980.

Lane, N. E. and Harris, S. D. Conversations with weapon systems: Crewstation applications of interactive voice technology. In Yearbook on Navy manpower, personnel & training research and development. Washington, D.C.: Office of the Chief of Naval Operations, in press.

Lea, W. A. Critical issues in airborne applications of speech recognition. Los Angeles, CA: Speech Communications Research Laboratory, 1980.

Lea, W. A. & Shoup, I. E. Review of the ARPA SUR project and survey of current technology in speech understanding. Los Angeles, CA: Speech Communications Research Laboratory, 1979.

Leitenberg, H., Rawson, R. A., & Mulick, J. A. Extinction and the reinforcement of alternative behavior. Journal of Comparative and Physiological Psychology, 1975, 88, 640-652.

Mandler, G. Mind and emotion. New York, NY: John Wiley & Sons, Inc., 1975.

McCauley, M. E. & Semple, C. A. Precision Approach Radar Training System (PARTS) training effectiveness evaluation, Preliminary Final Report NAVTRAEQUIPCEN 79-C-0042-1, Westlake Village, CA: Canyon Research Group, Inc., 1980.

McCormick, E. J. Human factors in engineering and design. New York, NY: McGraw-Hill Book Company, 1976.

Moshier, S. L., Osborn, R. R., Baker, J. M., & Baker, J. K. Dialog Systems automatic speech recognition capabilities present and future. In S. Harris (Ed.), Proceedings: Voice Interactive Systems: Applications and Payoffs, Dallas, TX, 1980. Reprinted by Naval Air Development Center, Warminster, PA, in press.

Rachlin, H. Behavior and learning. San Francisco, CA: W. H. Freeman and Company, 1976.

Reising, J. The crew-adaptive cockpit: Firefox, here we come. Proceedings of the Third Annual Conference on Digital Avionics Systems, Dallas, TX, 1979.

Skinner, B. F. Verbal behavior. New York, NY: Appleton-Century-Crofts, 1957.

Stokes, J. M., and Dow, L. Vocabulary Development for the Voice Recognition and Synthesis (VRAS) System. Technical Report 1400.05-A. Willow Grove, PA: Analytics, 1980.

Streib, M. I. & Preston, J. F. Voice recognition/synthesis for the Advanced Integrated Display System (AIDS), Technical Report 1343. Willow Grove, PA: Analytics, 1978.

Streib, M. I. & Stokes, J. M. Military applications of task oriented grammars, Technical Report 1400.10-B. Willow Grove, PA: Analytics, 1980.

Van Cott, H. P. and Kincade, R. G. (Eds.), Human engineering guide to equipment design. Washington, D.C.: American Institutes for Research, 1972.

Wagner, A. R. Conditioned frustration as a learnable drive. Journal of Experimental Psychology, 1963, 66, 142-148.

NAVTRAEQUIPCEN 80-D-0009-0155-1

DISTRIBUTION LIST

NAVAL TRAINING EQUIPMENT CENTER 25
ORLANDO, FL 32813

DEFENSE DOCUMENTATION CENTER 12
CAMERON STATION
ALEXANDRIA, VA 22310

All other addresses receive one copy

AEROSPACE PSYCHOLOGY DEPARTMENT
NAVAL AERO MEDICAL RESEARCH LAB (03)
NAVAL AIR STATION
PENSACOLA, FL 32508

CHIEF
ARI FIELD UNIT
P.O. BOX 2086
ATTN: LIBRARIAN
FT. BENNING, GA 31905

CHIEF
ARI FIELD UNIT
P. O. BOX 476
FT. RUCKER, AL 36362

ASD/AXA
ATTN: N.F. VIVIAN
WRIGHT-PATTERSON AFB, OH 45433

CHIEF OF NAVAL AIR TRAINING
ATTN: CODE 333
NAS
CORPUS CHRISTI, TX 78419

CHIEF OF NAVAL EDUCATION AND TRAINING
LIAISON OFFICE
HUMAN RESOURCE LABORATORY
FLYING TRAINING DIVISION
WILLIAMS AFB, AZ 85224

CHIEF OF NAVAL MATERIAL
MAT 0722
ATTN: ARNOLD I. RUBENSTEIN
WASHINGTON, DC 20360

CHIEF OF NAVAL OPERATIONS
OP-39T
WASHINGTON, DC 20350

CHIEF OF NAVAL OPERATIONS
OP-115
RESEARCH, DEVELOPMENT & STUDIES
ROOM G836
WASHINGTON, DC 20350

CHIEF OF NAVAL OPERATIONS
OP-596C
WASHINGTON, DC 20350

CHIEF OF NAVAL OPERATIONS
OP-987H
ATTN: DR. R. G. SMITH
WASHINGTON, DC 20350

CHIEF OF NAVAL RESEARCH
CODE 450
800 N. QUINCY STREET
ARLINGTON, VA 22217

CHIEF OF NAVAL RESERVE
CODE S-3311
NEW ORLEANS, LA 70146

COMMANDER
NAVAL AIR DEVELOPMENT CENTER
ATTN: CODE 6022
WARMINSTER, PA 18974

COMMANDER
NAVAL AIR FORCE
US PACIFIC FLEET (CODE 316)
NAS NORTH ISLAND
SAN DIEGO, CA 92135

COMMANDER
NAVAL WEAPONS CENTER
CODE 3154
ATTN: MR. BOB CURTIS
CHINA LAKE, CA 93555

COMMANDING OFFICER
FLEET ANTI-SUBMARINE WARFARE TRAINING
ATTN CODE 001
SAN DIEGO, CA 92147

COMMANDING OFFICER
NAVAL AEROSPACE MEDICAL RESEARCH LAB
SANDERS
DEPARTMENT OF PSYCHOLOGY
PENSACOLA, FL 32512

COMMANDING OFFICER
FLEET COMBAT TRAINING CENTER, PACIFIC
CODE 09A
SAN DIEGO, CA 92147

DIRECTOR
US ARMY RESEARCH 4 TECHNOLOGY LABORATORY
DAVDL-AS (ATTN: DR. R. S. DUNN)
207-S AMES RESEARCH CENTER
MOFFETT FIELD, CA 94035

DIRECTOR, NATIONAL SECURITY AGENCY
9800 SAVAGE ROAD
ATTN: MR. THOMAS W. PAGE, R54 FANX II
FT GEORGE G. MEADE, MD 20755

FEDERAL AVIATION ADMINISTRATION
AAC-950 ATT: AAC-954C
AERONAUTICAL CENTER, FLIGHT STANDARDS
OKLAHOMA CITY, OK 73101

HEADQUARTERS
AIR TRAINING COMMAND, XPTI
ATTN: MR. GOLDMAN
RANDOLPH AFB, TX 78148

HQAFSC/DLS
ANDREWS AFB, DC 20334

USAHEL/USAAVNC
ATTN: DRXHE-FR (MR. ARMSTRONG)
P. O. BOX 716
FT. RUCKER, AL 36362

US AIR FORCE HUMAN RESOURCES LAB
AFHRL-FT
FLYING TRAINING DIVISION
ATTN: DR. EDWARDS
WILLIAMS AFB, AZ 85224

US AIR FORCE HUMAN RESOURCES LAB
TSZ
BROOKS AFB, TX 78235

DR. DONALD W. CONNOLLY
RESEARCH PSYCHOLOGIST
FEDERAL AVIATION ADMINISTRATION
FAA NAFEC ANA-230
ATLANTIC CITY, NJ 08405

CDR P. M. CURRAN
NAVAL AIR DEVELOPMENT CENTER
TECHNOLOGY DEVELOPMENT BRANCH
CODE 602
WARMINSTER, PA 18974

DR. HENRY J. DEHAAN
US ARMY RESEARCH INSTITUTE
5001 EISENHOWER AVENUE
ALEXANDRIA, VA 22333

MR. HAROLD C. GLASS
US POSTAL LAB
11711 PARKLAWN DRIVE
ROCKVILLE, MD 20852

DR. HENRY M. HALFF
OFFICE OF NAVAL RESEARCH
CODE 458
ARLINGTON, VA 22217

MR. HAROLD A. KOTTMANN
ASD/YWE
WRIGHT PATTERSON AFB, OH 45433

CDR NORMAN E. LANE
NAVAL AIR DEVELOPMENT CENTER
CODE 6021
WARMINSTER, PA 18974

MR. ARTHUR W. LINDBERG
US ARMY AVIONICS R&D ACTIVITY
DAVAA-E
FT MONMOUTH, NJ 07703

CHARLES R. LUECK, JR
UNITED STATES POSTAL SERVICE
PROCESS CONTROL SYSTEMS TEST FACILITY
9201 EDGEWORTH DRIVE
WASHINGTON, DC 20027

MR. ERIC WERKOWITZ
AFFDL/FGR
WRIGHT PATTERSON AFB, OH 45433

MR. THOMAS J. MOORE
AFAMRL/BBA
WRIGHT PATTERSON AFB
DAYTON, OH 45433

MR. MELVYN C. MOY
NAVY PERSONNEL RESEARCH & DEVELOPMENT CTR
INFORMATION AND DECISION PROCESSES
CODE 305
SAN DIEGO, CA 92152

DR. JESSE ORLANSKY
INSTITUTE FOR DEFENSE ANALYSES
SCIENCE AND TECHNOLOGY DIVISION
400 ARMY-NAVY DRIVE
ARLINGTON, VA 22202

MR. GARY POOCK
NAVAL PG SCHOOL
CODE 55PK
MONTEREY, CA 93948

MR. ERNEST E. POOR
NAVAL AIR SYSTEMS COMMAND
AIR 413B
ROOM 336
WASHINGTON, DC 20361

JOHN SILVA
NAVAL OCEAN SYSTEMS CENTER
CODE 823
SAN DIEGO, CA 92152

MR. J. TRIMBLE
OFFICE OF NAVAL RESEARCH
CODE 240
800 N. QUINCY STREET
ARLINGTON, VA 22217

MR. LEAHMOND TYRE
FLEET MATERIAL SUPPORT OFFICE
CODE 9333
MECHANICSBURG, PA 17055

MR. HARRY A. WHITTED
CODE 8235
NAVAL OCEAN SYSTEMS CENTER
271 CATALINA BOULEVARD
SAN DIEGO, CA 92152

DR. TICE DE YOUNG
US ARMY ENGINEER TOPOGRAPHIC LABORATORIES
RESEARCH INSTITUTE
FT BELVOIR, VA 22060

